

# Chapter 14

## Measuring and Modeling Risk Using High-Frequency Data

Wolfgang Karl Härdle, N. Hautsch and U. Pigorsch

**Abstract** Measuring and modelling financial volatility is the key to derivative pricing, asset allocation and risk management. The recent availability of high-frequency data allows for refined methods in this field. In particular, more precise measures for the daily or lower frequency volatility can be obtained by summing over squared high-frequency re- turns. In turn, this so called realized volatility can be used for more accurate model evaluation and description of the dynamic and distributional structure of volatility. Moreover, non-parametric measures of systematic risk are attainable, that can straightforwardly be used to model the commonly observed time-variation in the betas. The discussion of these new measures and methods is accompanied by an empirical illustration using high-frequency data of the IBM incorporation and of the DJIA index.

### 14.1 Introduction

Volatility modelling is the key to the theory and practice of pricing financial products. Asset allocation and portfolio as well as risk management depend heavily on a correct modelling of the underlying(s). This insight has spurred extensive research in

---

W.K. Härdle (✉)

C.A.S.E.-Center of Applied Statistics and Economics, Humboldt-Universität zu Berlin,  
Unter den Linden 6, 10099 Berlin, Germany  
e-mail: haerdle@hu-berlin.de

W.K. Härdle

Ladislaus von Bortkiewicz Chair of Statistics, School of Business and Economics,  
Humboldt-Universität zu Berlin,  
Unter den Linden 6, 10099 Berlin, Germany

N. Hautsch

Department of Statistics and Operations Research, University of Vienna as well as Center for  
Financial Studies(CFS), Frankfurt, Germany  
e-mail: nikolaus.hautsch@univie.ac.at

U. Pigorsch

Schumpeter School of Business and Economics, University of Wuppertal, Wuppertal, Germany  
e-mail: pigorsch@uni-wuppertal.de

© Springer-Verlag GmbH Germany 2017

W.K. Härdle et al. (eds.), *Applied Quantitative Finance*, Statistics and Computing,  
DOI 10.1007/978-3-662-54486-0\_14

financial econometrics and mathematical finance. Stochastic volatility models with separate dynamic structure for the volatility process have been in the focus of the mathematical finance literature, see Heston (1993) and Bates (2000), while parametric GARCH-type models for the returns of the underlying(s) have been intensively analyzed in financial econometrics.

The validity of these models in practice though depends upon specific distributional properties or the knowledge of the exact (parametric) form of the volatility dynamics. Moreover, the evaluation of the predictive ability of volatility models is quite important in empirical applications. However, the latent character of the volatility poses a problem. To what measure should the volatility forecasts be compared to? Conventionally, the forecasts of daily volatility models, such as GARCH-type or stochastic volatility models, have been evaluated with respect to absolute or squared daily returns. In view of the excellent in-sample performance of these models, the forecasting performance, however, seems to be disappointing.

The availability of ultra-high-frequency data opens the door for a refined measurement of volatility and model evaluation. An often used and very flexible model for logarithmic prices of speculative assets is the (continuous time) stochastic volatility model:

$$dY_t = (\mu + \beta\sigma_t)dt + \sigma_t dW_t, \quad (14.1)$$

where  $\sigma_t^2$  is the instantaneous (spot) variance,  $\mu$  denotes the drift,  $\beta$  is the risk premium, and  $W_t$  defines the standard Wiener process. The object of interest is the amount of variation accumulated in a time interval  $\Delta$  (e.g., a day, week, month etc.). If  $n = 1, 2, \dots$  denotes a counter for the time intervals of interest, then the term

$$\sigma_n^2 = \int_{(n-1)\Delta}^{n\Delta} \sigma_t^2 dt \quad (14.2)$$

is called the actual volatility, see Barndorff-Nielsen and Shephard (2002b). The actual volatility is the quantity that reflects the market risk structure (scaled in  $\Delta$ ) and is the key element in pricing and portfolio allocation. Actual volatility (measured in scale  $\Delta$ ) is of course related to the integrated volatility:

$$V(t) = \int_0^t \sigma_s^2 ds \quad (14.3)$$

It is worth noting that there is a small notational confusion here: the mathematical finance literature would denote  $\sigma_t$  as “volatility” and  $\sigma_t^2$  as “variance”, see Nelson and Foster (1994). For example, an important result is that  $V(t)$  can be estimated from  $Y_t$  via the quadratic variation:

$$[Y_t]_M = \sum (Y_{t_j} - Y_{t_{j-1}})^2, \quad (14.4)$$

where  $t_0 = 0 < t_1 < \dots < t_M = t$  is a sequence of partition points and  $\sup_j |t_{j+1} - t_j| \rightarrow 0$ . Andersen and Bollerslev (1998) have shown that

$$[Y_t]_M \xrightarrow{P} V(t), \quad M \rightarrow \infty. \quad (14.5)$$

This observation leads us to consider in an interval  $\Delta$  with  $M$  observations

$$RV_n = \sum_{j=1}^M (Y_{t_j} - Y_{t_{j-1}})^2 \quad (14.6)$$

with  $t_j = \Delta\{(n-1) + j/M\}$ . Note that  $RV_n$  is a consistent estimator of  $\sigma_n^2$  and is called *realized volatility*. Barndorff-Nielsen and Shephard (2002b) point out that  $RV_n - \sigma_n^2$  is approximately mixed Gaussian and provide the asymptotic law of

$$\sqrt{M}(RV_n - \sigma_n^2). \quad (14.7)$$

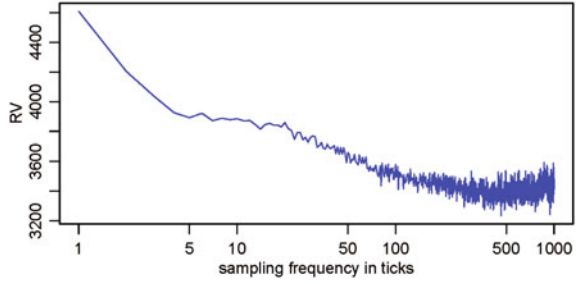
The realized volatility turns out to be very useful in the assessment of the validity of volatility models. For instance, reconciling evidence in favor of the forecast accuracy of GARCH-type models is observed when using realized volatility as a benchmark rather than daily squared returns. Moreover, the availability of the realized volatility measure initiated the development of a new and quite accurate class of volatility models. In particular, based on the ex-post observability of the realized volatility measure, volatility is now treated as an observed rather than a latent variable to which standard time series procedures can be applied.

The remainder of this chapter is structured as follows. We first discuss the practical problems encountered in the empirical construction of realized volatility which are due to the existence of market microstructure noise. Section 14.3 presents the stylized facts of realized volatility, while Sect. 14.4 reviews the most popular realized volatility models. Section 14.5 illustrates the usefulness of the realized volatility concept for measuring time-varying systematic risk within a conditional asset pricing model (CAPM).

## 14.2 Market Microstructure Effects

The consistency of the realized volatility estimator builds on the notion that prices are observed in continuous time and without measurement error. In practice, however, the sampling frequency is inevitably limited by the actual quotation or transaction frequency. Since high-frequency prices are subject to market microstructure noise, such as price-discreteness, bid-and-ask bounce effects, transaction costs etc., the true price is unobservable. Market microstructure effects induce a bias in the realized volatility measure, which can straightforwardly be illustrated in the following simple discrete-time setup. Assume that the logarithmic high-frequency prices are observed with noise, i.e.,

**Fig. 14.1** Volatility signature plot for IBM, 2001–2006. Average time between trades: 6.78 s. [XFGsignature](#)



$$Y_{t_j} = Y_{t_j}^* + \varepsilon_{t_j}, \tag{14.8}$$

where  $Y_{t_j}^*$  denotes the latent true price. Moreover, the microstructure noise  $\varepsilon_{t_j}$  is assumed to be iid distributed with mean zero and variance  $\eta^2$ , and is independent of the true return. Let  $r_{t_j}^*$  denote the efficient return, then the high-frequency continuously compounded returns

$$r_{t_j} = r_{t_j}^* + \varepsilon_{t_j} - \varepsilon_{t_{j-1}} \tag{14.9}$$

follow an MA(1) process. Such a return specification is well established in the market microstructure literature and is usually justified by the existence of the bid-ask bounce effect, see, e.g., Roll (1984). In this model, the realized volatility is given by

$$RV_n = \sum_{i=1}^M (r_{t_i}^*)^2 + 2 \sum_{j=1}^M r_{t_j}^* (\varepsilon_{t_j} - \varepsilon_{t_{j-1}}) + \sum_{j=1}^M (\varepsilon_{t_j} - \varepsilon_{t_{j-1}})^2. \tag{14.10}$$

with

$$E[RV_n] = E[RV_n^*] + 2M\eta^2. \tag{14.11}$$

If the sampling frequency goes to infinity, we know from the previous section that  $RV_n^*$  consistently estimates  $\sigma_n^2$  and, thus, the realized volatility based on the observed price process is a biased estimator of the actual volatility with bias term  $2M\eta^2$ . Obviously, for  $M \rightarrow \infty$ ,  $RV_n$  diverges.

This diverging behavior can also be observed empirically in so called volatility signature plots. Figure 14.1 shows the volatility signature for one stock of the IBM incorporation over the period ranging from January 2, 2001 to December 29, 2006. The plot depicts the average annualized realized volatility over the full sample period constructed at different frequencies measured in number of ticks (depicted in log scale). Obviously, the realized volatility is large at the very high frequency, but decays for lower frequencies and stabilizes around a sampling frequency of 300 ticks, which corresponds approximately to a 30 min sampling frequency, given that the average duration between two consecutive trades is around 6.78 s.

Thus, sampling at a lower frequency, such as every 10, 15 or 30 min, seems to alleviate the problem of market microstructure noise and has thus frequently been applied in the literature. This so-called *sparse sampling*, however, comes at the cost of a less precise estimate of the actual volatility. Alternative methods have been proposed to solve this bias-variance trade-off for the above simple noise assumption as well as for more general noise processes, allowing also for serial dependence in the noise and/or for dependence between the noise and the true price process, which is sometimes referred to as endogenous noise. A natural approach to reduce the market microstructure noise effect is to construct the realized volatility measure based on prefiltered high-frequency returns, using, e.g., an MA(1) model.

In the following we briefly present two more elaborate and under specific noise assumptions consistent procedures for estimating actual volatility. Both have been theoretically considered in several papers. The subsampling approach originally suggested by Zhang et al. (2005) builds on the idea of averaging over various realized volatilities constructed from different high-frequency subsamples. For the ease of exposition we focus again on one time period, e.g., one day, and denote the full grid of time points at which the  $M$  intradaily prices are observed by  $\mathcal{G}_t = \{t_0, \dots, t_M\}$ . The realized volatility that makes use of all observations in the full grid is denoted by  $RV_n^{(all)}$ . Moreover, the grid is partitioned into  $L$  nonoverlapping subgrids  $\mathcal{G}^{(l)}$ ,  $l = 1, \dots, L$ . A simple way for selecting such a subgrid may be the so-called regular allocation, in which the  $l$ -th subgrid is given by  $\mathcal{G}^{(l)} = \{t_{l-1}, t_{l-1+L}, \dots, t_{l-1+M_l L}\}$  for  $l = 1, \dots, L$ , and  $M_l$  denoting the number of observations in each subgrid. E.g., consider 5-min returns that can be measured at the time points 9:30, 9:35, 9:40, ..., and at the time points 9:31, 9:36, 9:41, ... and so forth. In analogy to the full grid, the realized volatility for subgrid  $l$ , denoted by  $RV_n^{(l)}$ , is constructed from all data points in subgrid  $l$ . Thus,  $RV_n^{(l)}$  is based on sparsely sampled data.

The actual volatility is then estimated by:

$$RV_n^{(ZMA)} = \frac{1}{L} \sum_{l=1}^L RV_n^{(l)} - \frac{\bar{M}}{M} RV_n^{(all)}, \quad (14.12)$$

where  $\bar{M} = \frac{1}{L} \sum_{l=1}^L M_l$ . The latter term on the right-hand side is included to bias-correct the averaging estimator  $\frac{1}{L} \sum_{l=1}^L RV_n^{(l)}$ . As the estimator (14.12) consists of a component based on sparsely sampled data and one based on the full grid of price observations, the estimator is also called the *two-timescales estimator*.

Given the similarity to the problem of estimating the long-run variance of a stationary time series in the presence of autocorrelation, it is not surprising that kernel-based methods have been developed for estimating the realized volatility. Most recently, Barndorff-Nielsen et al. (2008) proposed the flat-top realized kernel estimator

$$RV_n^{(BHLS)} = RV_n + \sum_{h=1}^{H^*} K\left(\frac{h-1}{H^*}\right) (\hat{\gamma}_h + \hat{\gamma}_{-h}) \quad (14.13)$$

with

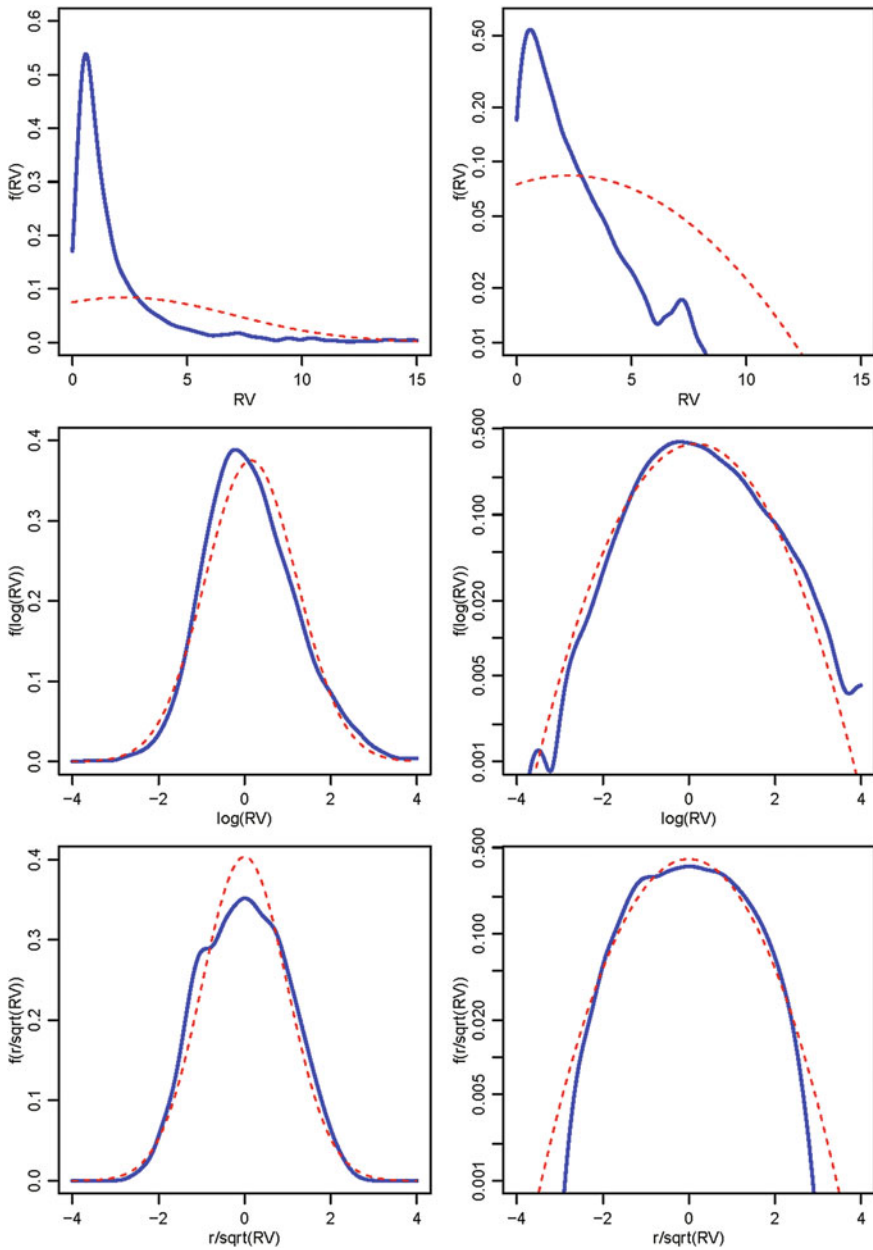
$$\hat{\gamma}_h = \frac{M}{M-h} \sum_{j=1}^M r_{t_j} r_{t_{j-h}}, \quad (14.14)$$

and  $K(0) = 1$ ,  $K(1) = 0$ . Obviously, the summation term on the righthand side is the realized kernel correction of the market microstructure noise. Zhou (1996), who was the first to consider realized kernels, proposed (14.13) with  $H = 1$ , while Hansen and Lunde (2006) allowed for general  $H$  but restricted  $K(x) = 1$ . Both of these estimators, however, have been shown to be inconsistent. Barndorff-Nielsen et al. (2008) instead propose several consistent realized kernel estimators with an optimally chosen  $H^*$ , such as the Tukey-Hanning kernel, i.e.  $K(x) = \{1 - \cos \pi(1-x)^2\}/2$ , which performs also very well in terms of efficiency as illustrated in a Monte Carlo analysis. They further show, that these realized kernel estimators are robust to market microstructure frictions that may induce endogenous and dependent noise terms.

### 14.3 Stylized Facts of Realized Volatility

Figure 14.2 shows kernel density estimates of the plain and logarithmic daily realized volatility in comparison to plots of a correspondingly fitted (log) normal distribution based on the IBM data, 2001–2006. The pictures in the top of Fig. 14.2 show the unconditional distribution of the (plain) realized volatility in contrast to a fitted normal distribution. As also confirmed by the corresponding descriptive statistics displayed by Table 14.1, we observe that realized volatility reveals severe right-skewness and excess kurtosis. This result might be surprising given that the realized volatility consists of the sum of squared intra-day returns and thus central limit theorems should apply. However, it is a common finding that intra-day returns are strongly serially dependent requiring significantly higher intra-day sampling frequencies to observe convergence to normality. In contrast, the unconditional distribution of the logarithmic realized volatility is well approximated by a normal distribution. The sample kurtosis is strongly reduced and is close to 3. Though slight right-skewness and deviations from normality in the tails of the distribution are still observed, the underlying distribution is remarkably close to that of a Gaussian distribution.

A common finding is that financial returns have fatter tails than the normal distribution and reveal significant excess kurtosis. Though GARCH models can explain excess kurtosis, they cannot completely capture these properties in real data. Consequently, (daily) returns standardized by GARCH-induced volatility, typically still show clear deviations from normality. However, a striking result in recent literature is that return series standardized by the square root of realized volatility,  $r_n/\sqrt{RV_n}$ , are quite close to normality. This result is illustrated by the plots in the bottom of



**Fig. 14.2** Kernel density estimates of the (logarithmic) realized volatility and of correspondingly standardized returns for IBM, 2001–2006. The *dotted line* depicts the density of the correspondingly fitted normal distribution. The *left* column depicts the kernel density estimates based on a log scale.

 [XFGkernelcom](http://XFGkernelcom)

**Table 14.1** Descriptive statistics of the realized volatility, log realized volatility and standardized returns, IBM stock, 2001–2006. LB (40) denotes the Ljung-Box statistic based on 40 lags. The last row gives an estimate of the order of fractional integration based on the Geweke and Porter-Hudak estimator

	$RV_n$	$\ln RV_n$	$r_n/\sqrt{RV_n}$
Mean	2.26	0.14	-0.000
Median	1.05	0.05	-0.013
Skewness	9.93	0.42	0.035
Variance	22.57	1.13	0.979
Kurtosis	150.47	3.43	2.349
1%-quantile	0.13	-2.03	-1.980
5%-quantile	0.24	-1.41	-1.558
95%-quantile	7.58	2.00	1.628
99%-quantile	17.66	2.87	2.141
LB(40)	2140.48	14213.07	39.780
p-value LB(40)	0.00	0.00	0.480
$\hat{d}$	0.38	0.62	-

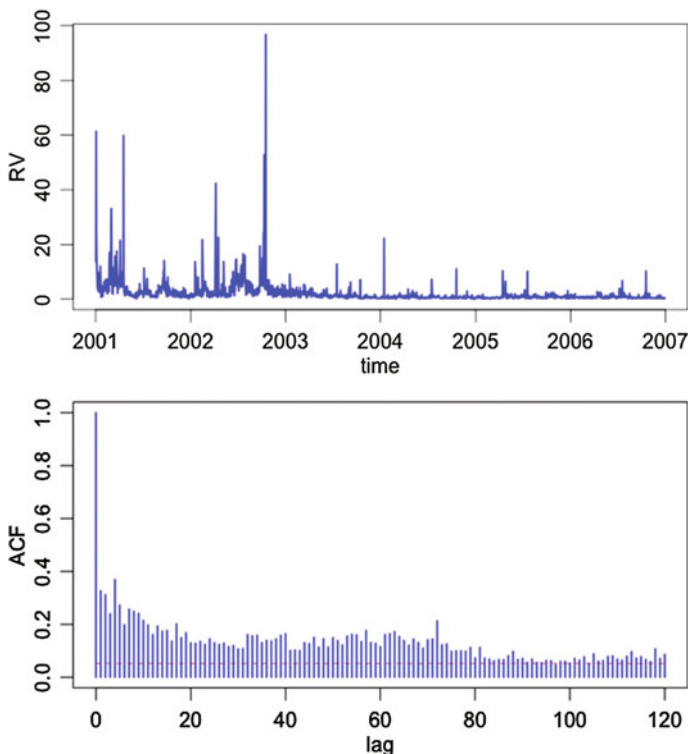
Fig. 14.2 and the descriptive statistics in Table 14.1. Though we observe deviations from normality for returns close to zero resulting in a kurtosis which is even below 3, the fit in the tails of the distribution is significantly better than that for plain log returns. Summarizing the empirical findings from Fig. 14.2, we can conclude that the unconditional distribution of daily returns is well described by a lognormal-normal mixture. This confirms the mixture-of-distribution hypothesis by Clark (1973) as well as the idea of the basic stochastic volatility model, where the log variance is modelled in terms of a Gaussian AR(1) process.


Figure 14.3 shows the evolution of daily realized volatility over the analyzed sample period and the implied sample autocorrelation functions (ACFs). As also shown by the corresponding Ljung-Box statistics in Table 14.1, the realized volatility is strongly positively autocorrelated with high persistence. This is particularly true for the logarithmic realized volatility. The plot shows that the ACF decays relatively slowly providing hints on the existence of long range dependence. Indeed, a common finding is that the realized volatility processes reveal long range dependence which is well captured by fractionally integrated processes. In particular, if  $RV_n$  is integrated of the order  $d \in (0, 0.5)$ , it can be shown that

$$\text{Var} \left[ \sum_{j=1}^h RV_{n+j} \right] \approx ch^{2d+1}, \quad (14.15)$$

with  $c$  denoting a constant. Then, plotting  $\ln \text{Var} \left[ \sum_{j=1}^h RV_{n+j} \right]$  against  $\ln h$  should result in a straight line with slope  $2d + 1$ . Most empirical studies strongly confirm





**Fig. 14.3** Time evolution and sample autocorrelation function of the realized volatility for IBM, 2001–2006.  XFGrvtsacf

this relationship and find values for  $d$  between 0.35 and 0.4 providing clear evidence for long range dependence. Estimating  $d$  using the Geweke and Porter-Hudak estimator, we obtain  $\hat{d} = 0.38$  for the series of realized volatilities and  $\hat{d} = 0.62$  for its logarithmic counterpart. Hence, for both series we find clear evidence for long range dependence. However, the persistence in logarithmic realized volatilities is remarkably high providing even hints on non-stationarity of the process.

Summarizing the most important empirical findings, we can conclude that the unconditional distributions of logarithmic realized volatility and of correspondingly standardized log returns are well approximated by normal distributions and that realized volatility itself follows a long memory process. These results suggest (Gaussian) ARFIMA models as valuable tools to model and to predict (log) realized volatility.

### 14.4 Realized Volatility Models

As illustrated above, realized volatility models should be able to capture the strong persistence in the sample autocorrelation function. While this seemingly long-memory pattern is widely acknowledged, there is still no consensus on the mechanism

generating it. One approach is to assume that the long memory is generated by a fractionally integrated process as originally introduced by Granger and Joyeux (1980) and Hosking (1981). In the GARCH literature this has led to the development of the fractionally integrated GARCH model as, e.g., proposed by Baillie et al. (1996). For realized volatility the use of a fractionally integrated autoregressive moving average (ARFIMA) process was advocated, for example, by Andersen et al. (2003). The ARFIMA ( $p, q$ ) model is given by

$$\phi(L)(1-L)^d(y_n - \mu) = \psi(L)u_n, \quad (14.16)$$

with  $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ ,  $\psi(L) = 1 + \psi_1 L + \dots + \psi_q L^q$ , and  $d$  denoting the fractional difference parameter. Moreover,  $u_n$  is usually assumed to be a Gaussian white noise process, and  $y_n$  denotes either the realized volatility (see Koopman et al. 2005) or its logarithmic transformation. Several extensions of the realized volatility ARFIMA model have been proposed, accounting, for example, for leverage effects (see Martens et al. 2004), for non-Gaussianity of (log) realized volatility or for time-variation in the volatility of realized volatility (see Corsi et al. 2008). Generally the empirical results show significant improvements in the point forecasts of volatility when using ARFIMA rather than GARCH-type models.

An alternative model for realized volatility has been suggested by Corsi (2009). The so-called heterogeneous autoregressive (HAR) model of realized volatility approximates the long-memory pattern by a sum of multi-period volatility components. The simulation results in Corsi (2009) show, that the HAR model can quite adequately reproduce the hyperbolic decay in the sample autocorrelation function of realized volatility even if the number of volatility components is small. For the HAR model, let the  $k$ -period realized volatility component be defined by the average of the single-period realized volatilities, i.e.,

$$RV_{n+1-k:n} = \frac{1}{k} \sum_{j=1}^k RV_{n-j}. \quad (14.17)$$

The HAR model with the so-defined daily, weekly and monthly realized volatility components, is given by

$$\begin{aligned} \log RV_n &= \alpha_0 + \alpha_d \log RV_{n-1} + \alpha_w \log RV_{n-5:n-1} \\ &\quad + \alpha_m \log RV_{n-21:n-1} + u_n, \end{aligned} \quad (14.18)$$

with  $u_n$  typically being a Gaussian white noise. The HAR model has become very popular due to its simplicity in estimation and its excellent in-sample fit and predictive ability (see e.g. Andersen et al. 2003; Corsi et al. 2008). Several extensions exist and deal, for example, with the inclusion of jump measures (see Andersen et al. 2003) or non-linear specifications based on neural networks (see Hillebrand and Medeiros 2007).

Alternative realized volatility models have been proposed in, e.g., Barndorff-Nielsen and Shephard (2002a), who consider a superposition of Ornstein Uhlenbeck processes, and in Deo et al. (2006), who specify a long-memory stochastic volatility model. A recent and comprehensive review on realized volatility models can also be found in McAleer and Medeiros (2008b).

## 14.5 Time-Varying Betas

So far, our discussion focused on the measurement and modeling of the volatility of a financial asset using high-frequency transaction data. From a pricing perspective, however, systematic risk is most important. In this section, we therefore discuss, how high-frequency information can be used for the evaluation and modeling of systematic risk. A common measure for the systematic risk is given by the so-called (market) beta, which represents the sensitivity of a financial asset to movements of the overall market. As the beta plays a crucial role in asset pricing, investment decisions, and the evaluation of the performance of asset managers, a precise estimate and forecast of betas is indispensable. While the unconditional capital asset pricing model implies a linear and stable relationship between the asset's return and the systematic risk factor, i.e., the return of the market, empirical results suggest that the beta is time-varying, see, for example, Bos and Newbold (1984), and Fabozzi and Francis (1978). Similar evidence has been found for multi-factor asset pricing models, where the factor loadings seem to be time-varying rather than constant. A large amount of research has therefore been devoted to conditional CAPM and APT models, which allow for time-varying factor loadings, see, for example, Dumas and Solnik (1995), Ferson and Harvey (1991), Ferson and Harvey (1993), and Ferson and Korajczyk (1995).

### 14.5.1 The Conditional CAPM

Below we consider the general form of the conditional CAPM. A similar discussion for multi-factor models can be found in Bollerslev and Zhang (2003). Assume that the continuously compounded return of a financial asset  $i$  from period  $n$  to  $n + 1$  is generated by the following process

$$r_{i;n+1} = \alpha_{i;n+1|n} + \beta_{i;n+1|n} r_{m;n+1} + u_{i;n+1}, \quad (14.19)$$

with  $r_{m;n+1}$  denoting the excess market return and  $\alpha_{n+1|n}$  denoting the intercept that may be time-varying conditional on the information set available at time  $n$ , as indicated by the subscript. The idiosyncratic risk  $u_{n+1}$  is serially uncorrelated,  $\mathbf{E}_n(u_{n+1}) = 0$ , but may exhibit conditionally time-varying variance. Note that  $\mathbf{E}_n(\cdot)$  denotes the expectation conditional on the information set available at time  $n$ . More-

over, we assume that  $E(r_{m;n+1}u_{n+1}) = 0$  for all  $n$ . The conditional beta coefficient of the CAPM regression (14.19) is defined as

$$\beta_{i;n+1|n} = \frac{\text{Cov}(r_{i;n+1}, r_{m;n+1})}{\text{Var}(r_{i;n+1})}. \quad (14.20)$$

Now, assume that lending and borrowing at a one-period risk-free rate  $r_{f;n}$  is possible. Then, the arbitrage-pricing theory implies that the conditional expectation of the next period's return at time  $n$  is given by

$$E_n(r_{i;n+1}) = r_{f;n} + \beta_{i;n+1|n} E_n(r_{m;n+1}). \quad (14.21)$$

Thus, the computation of the future return of asset  $i$  requires to specify how the beta coefficient evolves over time.

The most common approach to allow for time-varying betas is to re-run the CAPM regression in each period based on a sample of 3 or 5 years. We refer to this as the rolling regression (RR) method. More elaborate estimates of the beta can be obtained using the Kalman-filter, which builds on a statespace representation of the conditional CAPM or by specifying a dynamic model for the covariance matrix between the return of asset  $i$  and the market return.

### 14.5.2 Realized Betas

The evaluation of the in-sample fit and predictive ability of various beta models is also complicated by the unobservability of the true beta. Consequently, model comparisons are usually conducted in terms of implied pricing errors, i.e.,  $e_{i,n+1} = \widehat{r}_{i,n+1} - r_{i,n+1}$ , with  $\widehat{r}_{i,n+1} = r_{f;n} + \widehat{\beta}_{i;n+1|n} E_n(r_{m;n+1})$ . Owing to the discussion on the evaluation of volatility models, the question arises, whether high-frequency data may also be useful for the evaluation of competing beta estimates. The answer is a clear "yes". In fact, high-frequency based estimates of betas are quite informative for the dynamic behavior of systematic risk. The construction of so-called *realized betas* is straightforward and builds on realized covariance and realized volatility measures. In particular, denote the realized volatility of the market by  $RV_{m;n}$  and the realized covariance between the market and asset  $i$  by  $RCov_{m,i;n} = \sum_{j=1}^M r_{i,t_j} r_{m,t_j}$ , where  $r_{i,t_j}$  and  $r_{m,t_j}$  denote the  $j$ -th high-frequency return of the asset and the market, respectively, during day  $n$ . The realized beta is then defined as

$$\widehat{\beta}_{HF;i;n} = \frac{RCov_{m,i;n}}{RV_{m;n}}. \quad (14.22)$$

Barndorff-Nielsen and Shephard (2004) show that the realized beta converges almost surely for all  $n$  to the integrated beta over the time period from  $n - 1$  to  $n$ , i.e., the daily systematic risk associated with the market index. Note that the realized beta

can also be obtained from a simple regression of the high-frequency returns of asset  $i$  on the high-frequency returns of the market, see, e.g., Andersen et al. (2006). The preciseness of the realized beta estimator can easily be assessed by constructing the  $(1 - \alpha)$ -percent confidence intervals, which have been derived in Barndorff-Nielsen and Shephard (2004) and are given by

$$\widehat{\beta}_{HF;i;n} \pm z_{\alpha/2} \sqrt{\left(\sum_{j=1}^M r_{m,t_j}^2\right)^{-2} \widehat{g}_{i;n}}, \quad (14.23)$$

where  $z_{\alpha/2}$  denotes the  $(\alpha/2)$ -quantile of the standard normal distribution,

$$\widehat{g}_{i;n} = \sum_{j=1}^M x_{i;j}^2 - \sum_{j=1}^{M-1} x_{i;j} x_{i;j+1}, \quad (14.24)$$

and

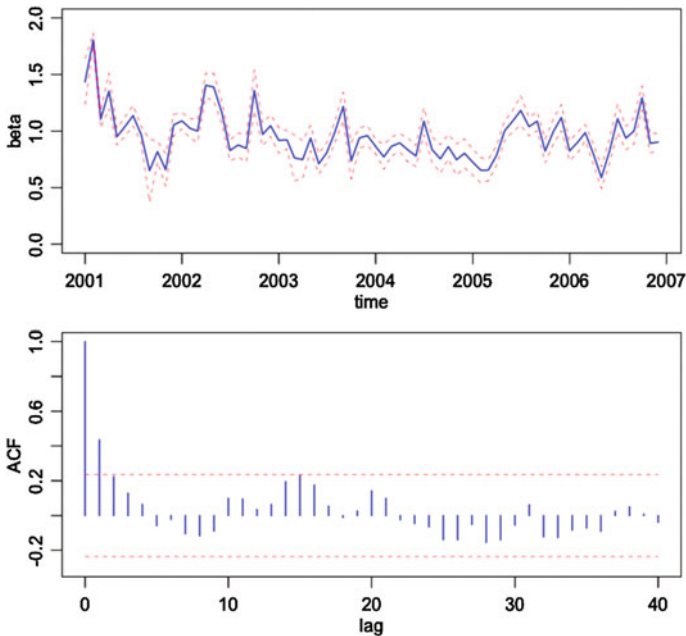
$$x_{i;j} = r_{i,t_j} r_{m,t_j} - \widehat{\beta}_{HF;i;n} r_{m,t_j}^2. \quad (14.25)$$

The upper panel in Fig. 14.4 presents the time-evolution of the monthly realized beta for IBM incorporation over the period ranging from 2001 to 2006. We use the Dow Jones Industrial Average Index as the market index and construct the realized betas using 30 min returns. The graph also shows the 95%-confidence intervals of the realized beta estimator. The time-varying nature of systematic risk emerges strikingly from the figure and provides once more evidence for the relevance of its inclusion in asset pricing models.

Interestingly, the sample autocorrelation function of the realized betas depicted in the lower panel of Fig. 14.4 indicates significant serial correlation over the short horizon. This dependency can be explored for the prediction of systematic risk. Bollerslev and Zhang (2003), for example, find that an autoregressive model for the realized betas outperforms the RR approach both in terms of forecast accuracy as well as in terms of pricing errors.

## 14.6 Summary

We review the usefulness of high-frequency data for measuring and modeling actual volatility at a lower frequency, such as a day. We present the realized volatility as an estimator of the actual volatility along with the practical problems arising in the implementation of this estimator. We show that market microstructure effects induce a bias to the realized volatility and we discuss several approaches for the alleviation of this problem. The realized volatility is a more precise estimator of the actual volatility than the conventionally used daily squared returns, and thus provides more



**Fig. 14.4** Time evolution and sample autocorrelation function of the realized volatility for IBM, 2001–2006. [XFGbetatsacf](#)

accurate information on the distributional and dynamic properties of volatility. This is important for many financial applications, such as asset pricing, portfolio allocation or risk management. As a consequence, several modeling approaches for realized volatility exist and have been shown to usually outperform traditional GARCH or stochastic volatility models, both in terms of in-sample as well as out-of-sample performance. We further demonstrate the usefulness of the realized variance and covariance estimator for measuring and modeling systematic risk. For the empirical examples provided in this chapter we use tick-by-tick transaction data of one stock of the IBM incorporation and of the DJIA index.

## References

- Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes standard volatility models do provide accurate forecasts. *International Economic Review*, 39, 885–905.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71, 579–625.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Wu, J. (2006). Realized beta: Persistence and predictability. In T. Fomby (Ed.), *Advances in econometrics: econometric analysis of economic and financial time series, volume B* (pp. 1–40).

- Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74, 3–30.
- Barndorff-Nielsen, O. E., & Shephard, N. (2002a). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, 64, 253–280.
- Barndorff-Nielsen, O. E., & Shephard, N. (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 4(5), 457–477.
- Barndorff-Nielsen, O. E., & Shephard, N. (2004). Econometric analysis of realized covariation: high frequency based covariance. *Regression, and Correlation in Financial Economics, Econometrica*, 72, 885–925.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, forthcoming.
- Bates, D. S. (2000). Post-'87 crash fears in the S&P 500 futures option. *Journal of Econometrics*, 94(1–2), 181–238.
- Bollerslev, T., & Zhang, B. Y. B. (2003). Measuring and modeling systematic risk in factor pricing models using high-frequency data. *Journal of Empirical Finance*, 10, 533–558.
- Bos, T., & Newbold, P. (1984). An empirical investigation of the possibility of stochastic systematic risk in the market model. *Journal of Business*, 57, 35–41.
- Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41, 135–156.
- Corsi, F. (2009). A simple long memory model of realized volatility. *Journal of Financial Econometrics*, 7, 174–196.
- Corsi, F., Mittnik, S., Pigorsch, C., & Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27, 46–78.
- Deo, R., Hurvich, C., & Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics*, 131(1–2), 29–58.
- Dumas, B., & Solnik, B. (1995). The world price of exchange rate risk. *Journal of Finance*, 50, 445–480.
- Fabozzi, F. J., & Francis, J. C. (1978). Beta as a random coefficient. *Journal of Financial and Quantitative Analysis*, 13, 101–116.
- Ferson, W. E., & Harvey, C. R. (1993). The risk and predictability of international equity returns. *Review of Financial Studies*, 6, 527–566.
- Ferson, W. E., & Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, 99, 385–415.
- Ferson, W. E., & Korajczyk, R. A. (1995). Do arbitrage pricing models explain the predictability of stock returns? *Journal of Business*, 68, 309–349.
- Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227–238.
- Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis*, 1, 15–30.
- Granger, C. W. J., & Teräsvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economic Letters*, 62, 161–165.
- Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24, 127–161.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2), 327–343.
- Hillebrand, E., & Medeiros, M. (2007). *Forecasting realized volatility models: The benefits of bagging and non-linear specifications*. Louisiana State University, Working Paper.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68, 165–176.

- Koopman, S. J., Jungbacker, B., & Hall, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12(3), 445–475.
- Martens, M., & Zein, J. (2004). Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets*, 11, 1005–1028.
- Martens, M., van Dijk, D., & dePooter, M. (2004). *Modeling and forecasting S&P 500 volatility: Long memory, structural breaks and nonlinearity*. Erasmus University Rotterdam, Working Paper.
- McAleer, M., & Medeiros, M. (2008a). A multiple regime smooth transition heterogenous autoregressive model for long memory and asymmetries. *Journal of Econometrics*, 147, 104–119.
- McAleer, M., & Medeiros, M. (2008b). Realized volatility: A review. *Econometric Reviews*, 26, 10–45.
- Müller, U. A., Dacorogna, M. M., Dav, R. D., Olsen, R. B., Pictet, O. V., & von Weizsäcker, J. E. (1997). Volatilities of different time resolutions-analyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2–3), 213–239.
- Nelson, D. B., & Foster, D. P. (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica*, 62, 1–41.
- Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39, 1127–1139.
- Zhang, L., Mykland, P. A., & Ait-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472), 1394–1411.
- Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14, 45–52.